Received: 5 November 2019 | Revised: 10 November 2020 | Accepted: 7 December 2020

IET Computer Vision

DOI: 10.1049/cvi2.12028

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Coconut trees detection and segmentation in aerial imagery using mask region-based convolution neural network

Muhammad Shakaib Iqbal[1] | **Hazrat Ali**[1] | **Son N. Tran**[2] | **Talha Iqbal**[3]

[1]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad, Pakistan

[2]Department of Information and Communication Technology, University of Tasmania, Australia

[3]Smart Sensor Lab, School of Medicine, National University of Ireland, Galway, Ireland

Correspondence

Hazrat Ali, Department of Electrical and Computer Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan.
Email: hazratali@cuiatd.edu.pk

Abstract

Food resources face severe damages under extraordinary situations of catastrophes such as earthquakes, cyclones, and tsunamis. Under such scenarios, speedy assessment of food resources from agricultural land is critical as it supports aid activity in the disaster-hit areas. In this article, a deep learning approach was presented for the detection and segmentation of coconut trees in aerial imagery provided through the AI competition organised by the World Bank in collaboration with *OpenAerialMap* and *WeRobotics*. Masked Region-based Convolution Neural Network (Mask R-CNN) approach was used for identification and segmentation of coconut trees. For the segmentation task, Mask R-CNN model with ResNet50 and ResNet101 based architectures was used. Several experiments with different configuration parameters were performed and the best configuration for the detection of coconut trees with more than 90% confidence factor was reported. For the purpose of evaluation, Microsoft COCO dataset evaluation metric namely mean average precision (mAP) was used.An overall 91% mean average precision for coconut trees' detection was achieved.

1 | INTRODUCTION

Natural disasters in the Kingdom of Tonga (South Pacific) are an unfortunate global reality. Their consequences can be damaging for the south Pacific population who heavily depend on the local agriculture as a primary food source [1]. As per the 2015 statement of the Secretary-General of the UN on the 'Implementation of the International Strategy for Disaster Reduction', approximately 1.5 trillion USD losses have incurred as a direct consequence of the natural catastrophes around the world[1]. The rate of recurrence as well as the magnitude of the severity of these disasters are increasing. Hence, there is a great demand to reinforce food security mechanisms and make appropriate assessments of the damages caused [2].

When cyclones strike, recognising the area of damage is crucial for effective humanitarian response and securing undamaged food sources like the coconut trees. The World Bank seeks qualified teams to develop machine- learning-based methods to automate the assessment of aerial imagery and to classify and locate the standing trees such as coconut trees within the aerial snapshot [3]. Manual aerial image classification is a resource and skill−intensive task and requires a lot of time. More importantly, manual aerial image classification is not typically risk-free in disaster-hit regions.

OpenAerialMap, World Bank, and WeRobotics have collaboratively launched an open machine-learning challenge to speed up the classification and analysis of high-resolution aerial imagery before and after humanitarian disaster [4]. The idea is to explore and develop machine-learning solutions for the classification of various features of interest in aerial imagery obtained through UAV. The features thus obtained can then be utilised for object detection and classification to help in the assessment of damages caused. One of the tasks in the challenge is to build a model for coconut trees' detection. In this task, a spatial high-resolution image (about 8 cm/pixel) was given, which covers 50 km^2 Area of Interest (AOI) of the kingdom of Tonga in the south Pacific region. The imagery was taken during October 2017, which is recent. Along with the aerial image, the shape (.shp) files were provided to recognise the geometric locations and classes of the targets (objects of interest) like roads and trees. Data for relevant features have been labelled by the group of volunteers from the humanitarian OpenStreetMap (OSM) community[2].

Object detection in aerial imagery is an interesting task and has attracted the computer vision and machine-learning research community [5–7]. Typically, these approaches use Convolutional Neural Networks (CNN) for object detection. The prevailing work has mostly been done on road detection and vehicle detection [5–7]. In this article, a framework was developed to detect and locate coconut trees. More specifically, it addresses the task of coconut trees' classification and localization. With the help of experimental results, the use of mask R-CNN can be demonstrated to detect coconut trees within the images. This is challenging as some of these images include mislabelled and missing ground truth entries. Besides, generic shapes that have different objects are difficult to be differentiated in the aerial image. Finally, there are many small objects occupied by densely concentrated regions in the aerial image.

This contributions of this article are as follows:

- A framework was presented for automatic detection and localization of coconut trees within given aerial imagery. The framework is able to detect each individual coconut tree with a high confidence factor and provide a segmented mask.
- The proposed framework for coconut trees detection provides a baseline approach, which can be easily extended to detect and identify other types of trees.
- Agriculture resource management is a labour-intense and high-risk job. The proposed approach provides a low-cost solution for agriculture resource management and measurement of the impact of disasters on natural food resources while reducing the risk factors for human operators.

The rest of the article is organised as follows: In Section 2 a brief overview of deep learning techniques for objection detection is provided. In Section 3, a detailed description of the dataset, methods, and training mechanism used in this work is provided. The results on coconut trees prediction with the help of figures in Section 4, and also provide a discussion on the model and results obtained are presented. Finally, The conclusion of this article is provided in Section 5.

2 | RELATED WORK

Over the last few decades, satellite imagery has been often used in a diversified range of applications ranging from forestry [8] to agriculture [9, 10], target detection [11] and regional planning to warfare [12]. Satellite imagery has also been broadly employed to monitor natural disasters and various other adverse incidents to investigate their impact on the environment.

Deep learning modifies the traditional machine learning by addition of more 'depth' in the model and transforming the information through several layers and non-linearity functions. This provides hierarchical data representation through abstraction of many levels [13]. Deep learning extracts useful features from raw data, with features from high levels of the hierarchy shaped through a combination of low level features [14]. The huge parallelisation possible in deep learning models enables to develop highly complex models for learning complex features and performing extremely well on many AI tasks [15]. So, deep learning models can enhance categorisation efficiency or minimise error in regression problems, given adequate large data is available for a specific domain task.

The large capacity of models and highly hierarchical structure performs very well particularly on the prediction and classification tasks, being adaptable and flexible for a broad range of highly complex problems [15]. Although deep learning has got fame in various applications, on coping with raster-based information (such as pictures, videos), it can be applied to an array of different types of information, that is speech, audio and natural language, and other data types like population information [16], continuous data such as weather data [17] and soil chemistry [17]. The vital role of utilising deep learning in the processing of images is the reduced need for feature engineering. In the past, conventional methods for image classification tasks were typically based on manual hand-engineered features. However, feature engineering is a time-consuming, costly, and complex method that needs to be changed whenever the data-set or the problem changes. Thus, feature engineering involves a costly effort, which is based on the expert's ability and may not generalise well [18]. Alternatively, a deep learning model does not rely on feature engineering and rather learn features through representation learning.

The region-based convolutional neural network (R-CNN) has proved to be very successful for segmentation tasks [19, 20]. In R-CNN, a selective search technique is applied to detect region proposals within the input image. Region proposals structure the features vector that is given to multiple classifiers to represent a distribution of class variables and also to a regression model to refine bounding boxes of regions of a proposal. Fast R-CNN and Faster R-CNN proposed in [21, 22], respectively, accelerate the detection procedure by first applying a deep CNN to the input image and then extract features map and simply swap the selective-search by the Region Proposal Network (RPN) to create region proposals, predicting bounding boxes and classes of the objects. The extension of Faster R-CNN to Mask R-CNN [23] puts a parallel branch to object detection to predict object masks with very small overhead. Mask R-CNN outperformed top models in the 2017 COCO competition in segmentation, object detection, and bounding-box detection.

Hence, Mask R-CNN is the preferred model. Recently reported attempts on object detection and image detection include [24, 25]. The work in [24] uses the U-Net architecture with a ResNet decoder and the work in [25] produces geometry preserved masking for a better fit on the object's boundaries. The selection of Mask R-CNN in this work was made based upon two considerations: 1) The model outperformed top models on the most recent COCO competition for object detection and 2) the underlying software tools and tensor-flow libraries make the import of the pre-trained weights relatively easier making the implementation more convenient.

3 | PROPOSED METHODOLOGY

The overall pipeline is summarised in Figure 1. Briefly, the data is pre-processed and divided into training and test sets. Since pre-trained weights are utilised, it is not necessary to train the whole neural network. The final layers are trained (bounding box heads/classification) and the configuration settings were selected with the minimum validation error. At last, the overall performance was evaluated using the previously unseen test data. In the following sub-sections, these phases are discussed in more detail. To generalise this model, the following steps have been taken:

- A stronger weight decay is used that is, the L2 regularisation.
- An optimal learning rate of 0.001 (to avoid converging to local minima) after trying out different values among {0.0001, 0.0003, 0.001, 0.003, 0.01}.
- An adaptive optimiser named as SWATS, an approach to switch optimization from Adam to Stochastic gradient descent and thus achieving better generalisation was used [26].

3.1 | Data processing

As discussed in the introduction section, one single high-resolution aerial snapshot and a shape file (GIS file that stores geometric locations as well as attributes of geometric features, such as points, lines, and polygons) are provided. The objective is to combine and convert both data sources into training, validation, and test sets appropriate for the object detection model. GDAL is an open-source library to deal with raster and vector geospatial data types[3]. Fiona is a very popular python library for writing and reading geospatial records[4]. Fiona is used to read the shape file into a JSON file format where the geographical data is given and it is readily available for data processing. After that, the positions of the items of interests are extracted, such as coconut trees by looking at the tags offered for every geo JSON object in the shape file. The positions are given in the latitudinal and longitudinal coordinate system. To map these locations on to the very high-resolution aerial image, the actual image is converted into the latitude-longitude coordinate system by using GDAL tools. Furthermore, the latitude and longitude coordinates are mapped which are taken out from the shape file into the image-pixels by using the geographical metadata in the high-resolution aerial image. At the end of the procedure, an image combined with the pixels is obtained where objects of interests (coconuts trees) are located.

The input to the Mask R-CNN framework is the set of annotated train image tiles of size 1000 × 1000 pixels. The actual image is subdivided into patches of the dimensions 1000 × 1000. Then,70 such tiles are taken and every single tile is manually annotated by positioning the coconut trees and then drawing polygons around them. It really is a time-consuming procedure, but it is believed that a correctly annotated dataset is vital for training a model with high prediction accuracy. VGG Image Annotator[5] is used to promptly annotate 70 image tiles. Every single annotation consists of JSON file format and keeps the positions of all polygons along with their tags. There are approximately 40 and 60 objects in each tile, therefore it is concluded that the training dataset is substantial. A good example of an image tile annotated with the VGG Image Annotator is displayed in Figure 2. The training and validation losses are shown in Figure 3 and Figure 4 respectively.

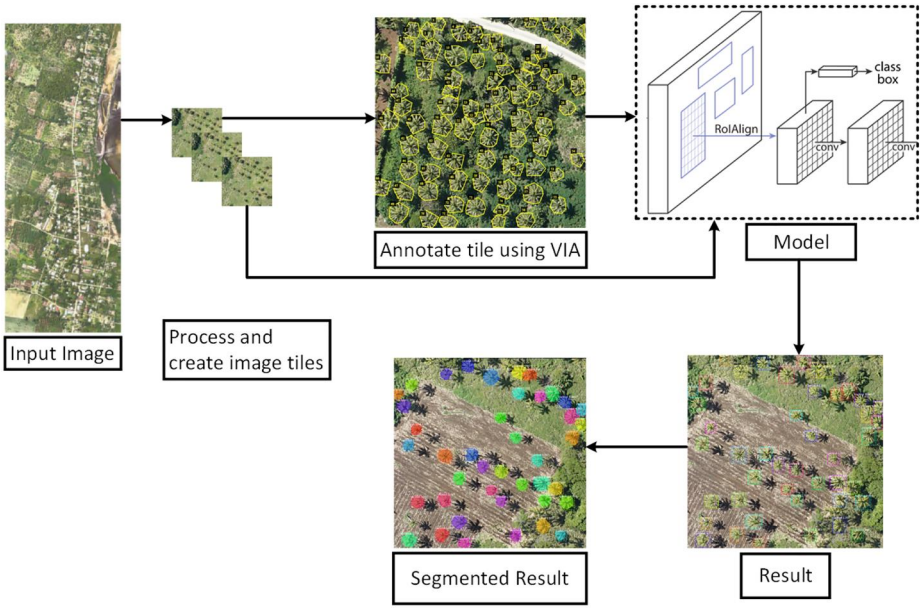

FIGURE 1 Proposed Model Pipeline (From left to right) Original snapshot is categorised into tiles divided into training and test examples. Annotated images are fed to the pre-trained model. Final layers of the model are fine-tuned on our own dataset after that model is ready for detection

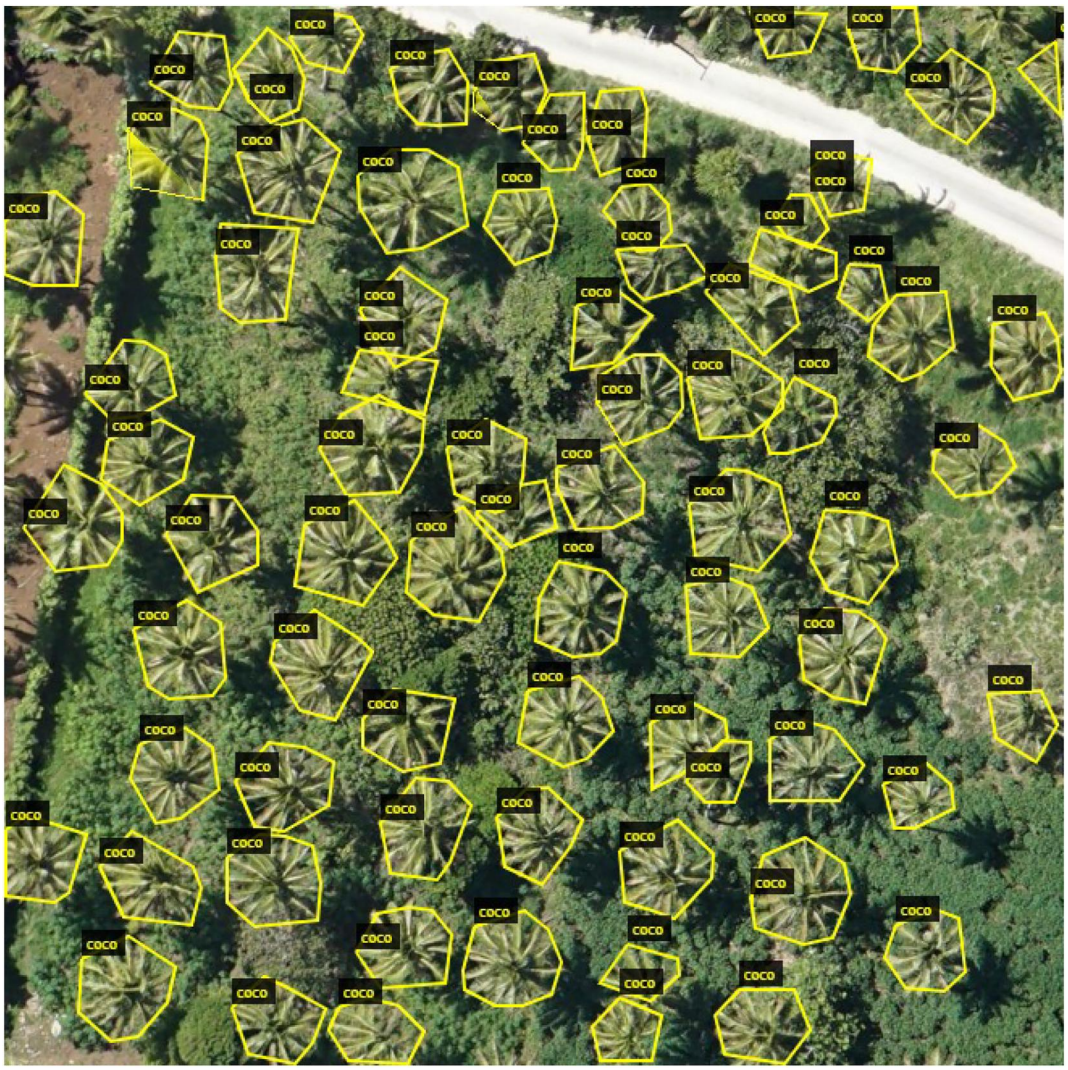

FIGURE 2 Sample of training data. The image annotation is done using VGG image annotator. Only coconut trees are annotated

It is very important to note, that during the annotation a lot of discrepancies are found. For example, a few coconut trees were mislabelled (such as coconut trees labelled as banana trees, dark areas and shadows labelled as trees) and several trees were not marked at all.

3.2 | Deep learning architecture and training

As discussed in the model selection phase, the foundation of this approach is based on the Mask R-CNN implementation. The ResNet101, a deep residual neural network with 101 layers, is the backbone architecture that extracts feature maps from the input image [27]. Residual networks enables to efficiently train deep neural networks simply by introducing skip connections, in which weights coming from previous layers are copied into a more deep layer. It requires an image of $1000 \times 1000 \times 3$ and then outputs feature map of dimension $32 \times 32 \times 2048$. These features are moved to an RPN for training regression/classification of object classes and generation of bounding boxes.

The training procedure was initiated by downloading a model pre-trained on the Microsoft COCO dataset, one of the most widely used datasets for object detection and segmentation. In principle, the setting of the earlier layers are not changed, but few RPN parameters are modified as it is aimed to train the final layers of the model (referred to as regression and classification heads). To accelerate the training process, although it is aimed to attain high accuracy, the minimum region proposal confidence is specified to 0.9, which means only regions with more than 90% confidence of potentially containing trees are considered. The confidence score of 0.9 is selected after an empirical evaluation of different possible values. Selecting a value less than 0.9 causes classifiers to incorrectly detect shadows and other trees like objects in the image as coconut trees (more false positives), as shown in Figure 5. On the other hand, a confidence value greater than 0.9 results in missing out some of the coconut trees, such as trees behind light clouds, which are not detected properly

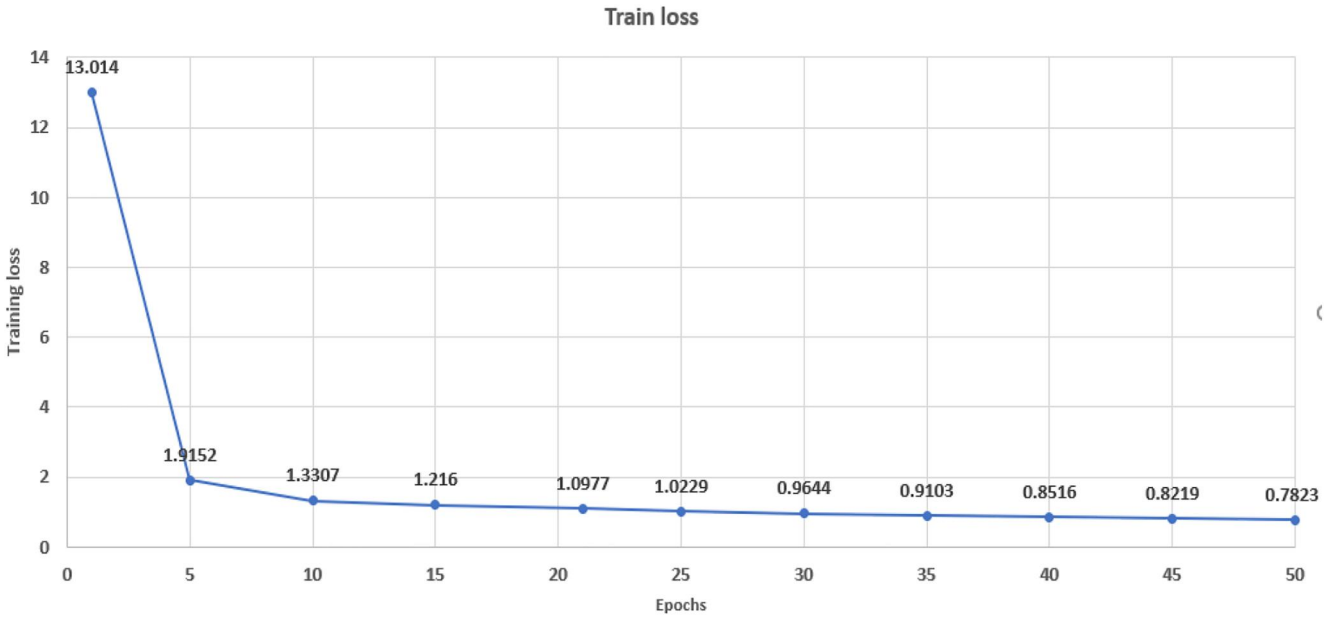

FIGURE 3 Training loss from epoch 1 to 50. Changes in the training loss after 21 epochs are not significant

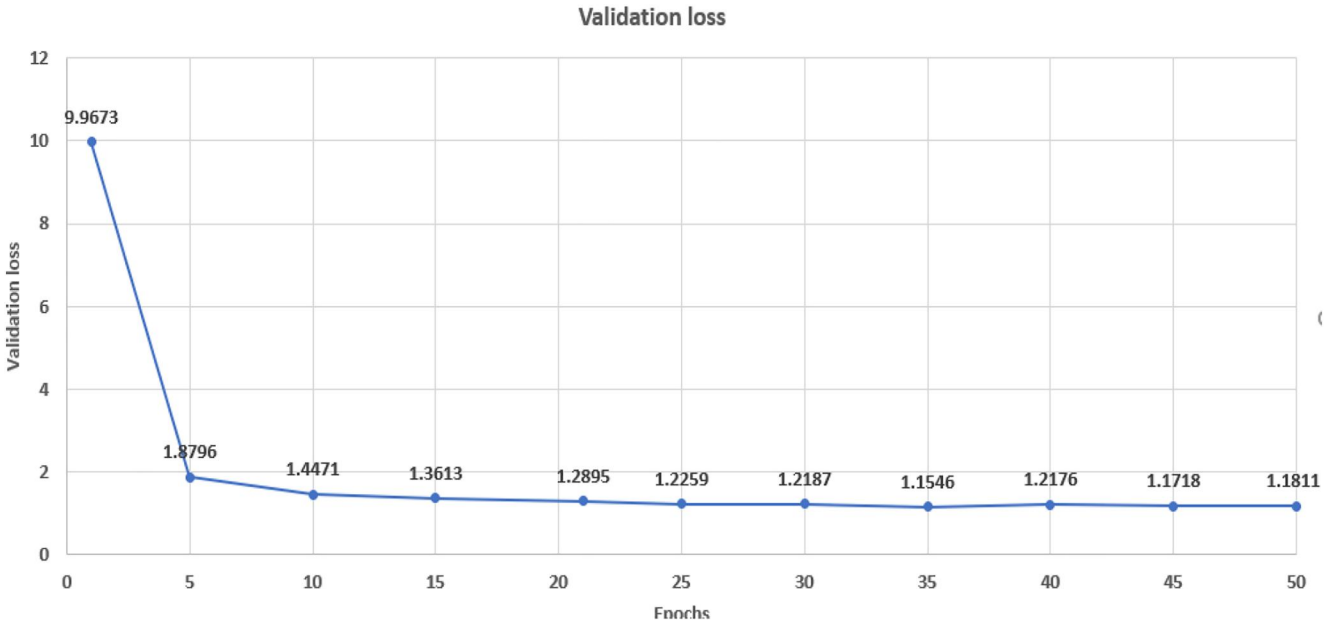

FIGURE 4 Validation loss from epoch 1 to 50. The validation loss changes very slowly after 21 epochs and the change is not significant

(more false negatives), eventually resulting in a lower detection accuracy, as shown in Figure 6. Furthermore, since in RGB aerial images coconut trees are expected to have approximately similar aspect ratios and sizes, anchor scales are set between 10 130. The learning rate is set to 0.001 while using the weight decay of 0.0001. The available data (70 tiles) was divided into training/validation/test sets of 50/10/10 image tiles and several experiments were performed by changing the number of steps, the number of epochs, and the number of maximum possible Regions of Interest (ROIs).

According to the validation scores of multiple experiments that are executed, the configuration having the best performance was selected: train(Q) for 21 epochs, consist of 100 steps each, with the maximum number of ROIs being 110 (this is a good choice since a maximum of 70 trees per image tile was obtained as observed during the annotation phase).

4 | RESULTS AND DISCUSSIONS

The weights of this model is applied and trained for 21 epochs to detect coconut trees on the test-set that consists of 10 images[6]. After performing several experiments, best configuration settings are shown in Table 1. The prediction results on some of the images are indicated in Figure 7 and Figure 8 The segmentation results are shown in Figure 9

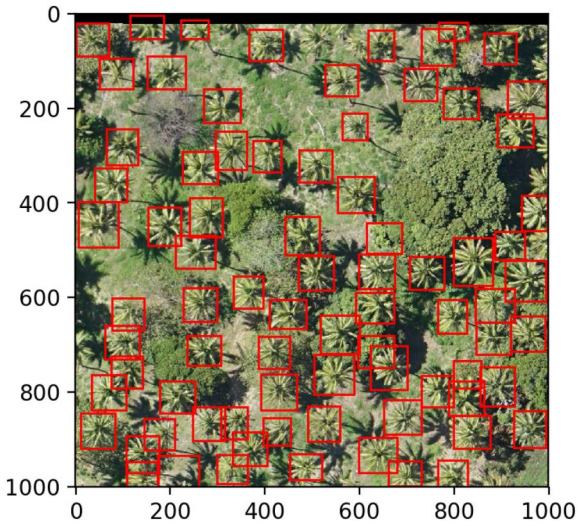

FIGURE 5 Bounding box predictions using ResNet101. All coconut trees prediction confidences are above 70%. The algorithm has incorrectly detected some coconut trees as threshold confidence is set low

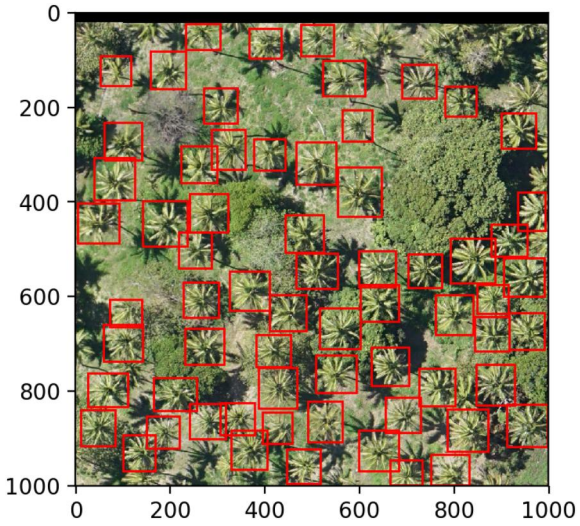

FIGURE 6 Bounding box predictions using ResNet101. All coconut trees' prediction confidences are above 95%. The algorithm has missed out some coconut trees as threshold confidence is set high

TABLE 1 Configuration settings

Backbone	ResNet101
Batch size	1
Detection min confidence	0.9
Detection max instances	100
Learning momentum	0.9
Learning rate	0.001
Steps per epoch	100
Train rois per image	110
Validation steps	50
Weight decay	0.0001
Epochs	50

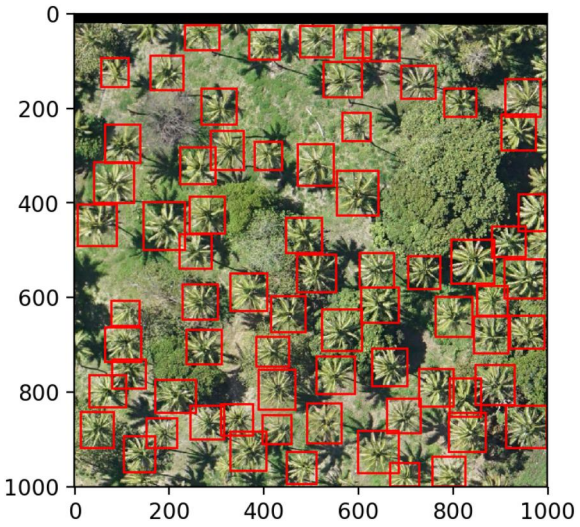

FIGURE 7 Bounding box predictions using ResNet50. Many but not all coconut trees are detected. Few shadows are detected as coconut trees

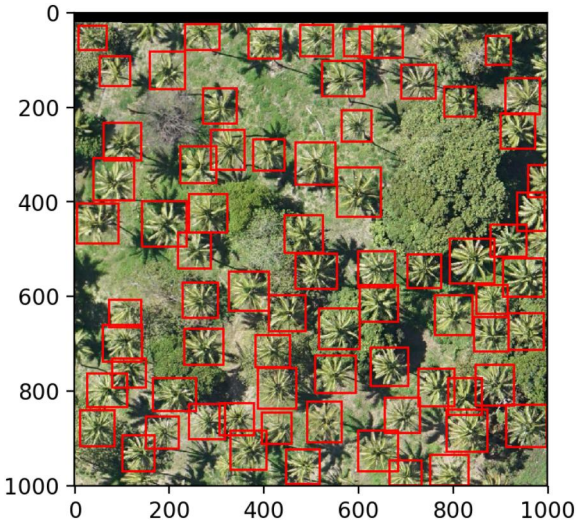

FIGURE 8 Bounding box predictions using ResNet101. ResNet101 results are better than ResNet50 on the Bounding box predictions. All coconut trees prediction confidences are above 90%

The overall training time for 5 and 10 batch size was less compared with batch size one, but training and validation loss of batch size of 5 and 10 is greater than batch size = 1. So, batch size = 1 was selected for this experiment, as shown in Figures 10 and 11. A comparison table on training and validation loss for the different choices of batch sizes is shown in Table 3. Optimal learning rate of 0.001 (to avoid converging to local minima) after trying out different values among [0.0001, 0.0003, 0.001, 0.003, 0.01].

4.1 | Configuration settings

After performing several experiments, the best configuration settings which is set for our project are shown in Table I.

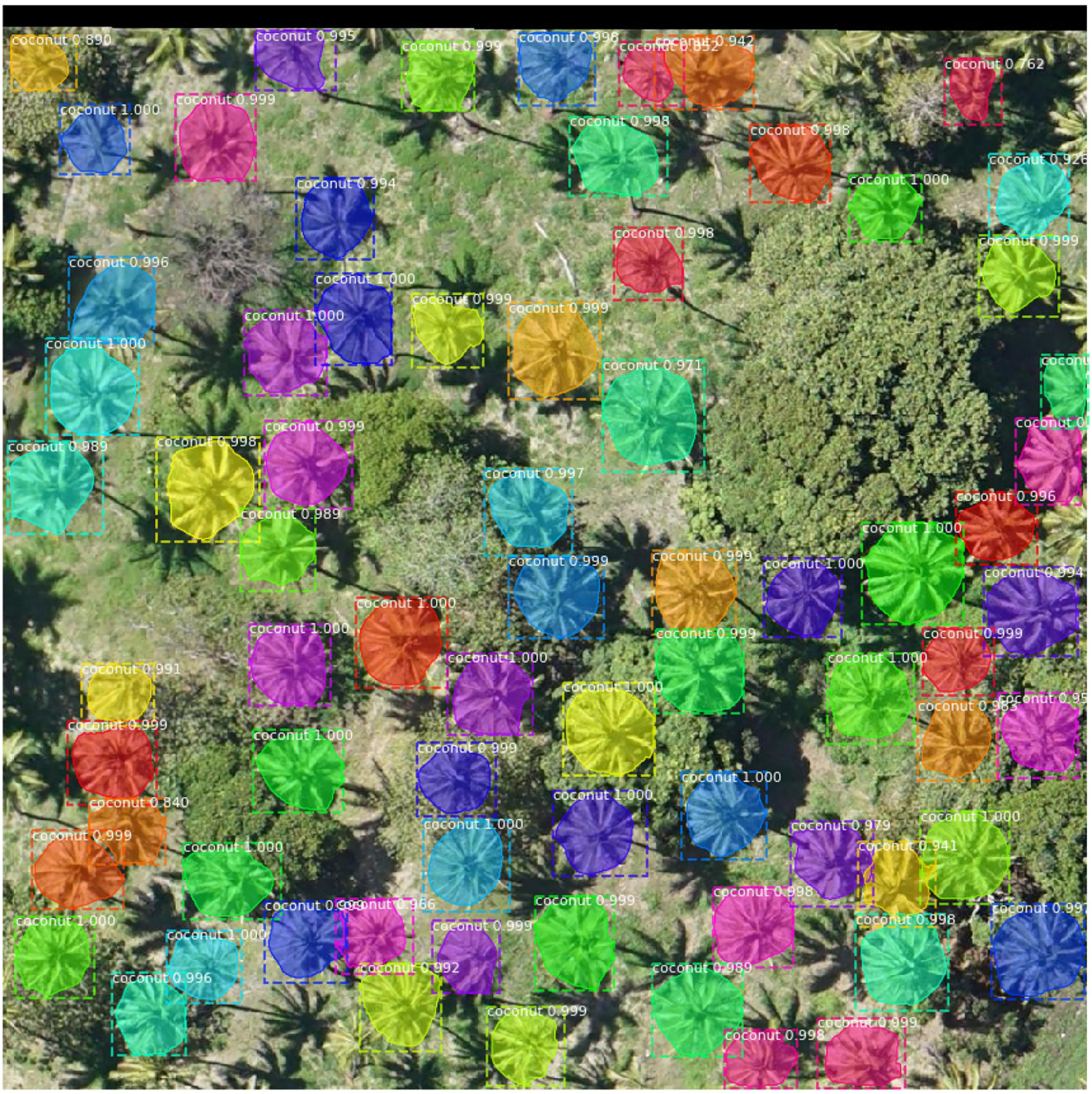

FIGURE 9 Coconut trees prediction. Segmentation results using ResNet101 as backbone architecture. (Best seen in colour)

5 | DISCUSSION

It was determined that all coconut trees were detected with a considerably high confidence factor (which is >90%). Classification accuracy (CA) of 96% was achieved with ResNet50% and 98% CA using ResNet101. For a more formal evaluation, mAP metric, a commonly used metric for performance evaluation of object detection was selected. The mAP is a mean of average precision, where not only the identification number but also the order of the correct predictions is evaluated. The highest mAP value achieved was 0.88 for ResNet50 and 0.91 for ResNet101. Detection results are visualised in Figures 5 and 6. The mAP curves are shown accordingly in Figure 12 and Figure 13. F1 Score is 0.89 for ResNet50 and 0.92 for ResNet101. Evaluation metrics are shown in Table 2. The processing time of this approach is one minute for 1000 × 1000 image. Table 3 summarises the training and validation losses with different batch sizes used for training the model. The overall training and validation losses of batch size one and batch size 10 are almost similar despite batch size 10 having less training time. The proposed algorithm is compared with other state-of-the-art techniques used for tree detection and segmentation, in the next subsection.

FIGURE 10 Graph of train loss, validation loss of overall experiment reported up-till 50 epochs using batch size 5

FIGURE 11 Graph of train loss, validation loss of overall experiment reported up-till 50 epochs using batch size 10

5.1 | Comparison with other techniques

Some excellent works reported on trees' detection from high-resolution aerial imagery involve different datasets, pre-processing methods, models, parameters, and metrics. A direct comparison is not made as the datasets used or the tasks performed in these approaches are different. However, it is still useful to provide a summary of the results of these

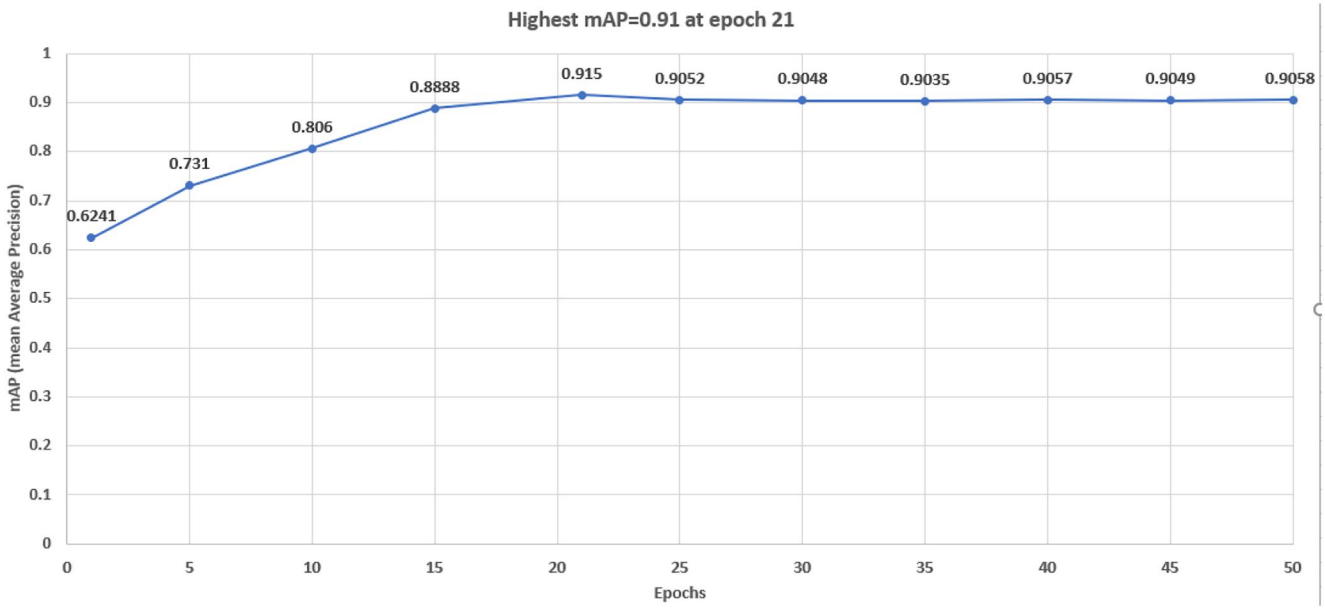

FIGURE 12 Highest mean average precision (mAP) achieved is at epoch number 21 using ResNet101 as the backbone network in Mask R-CNN

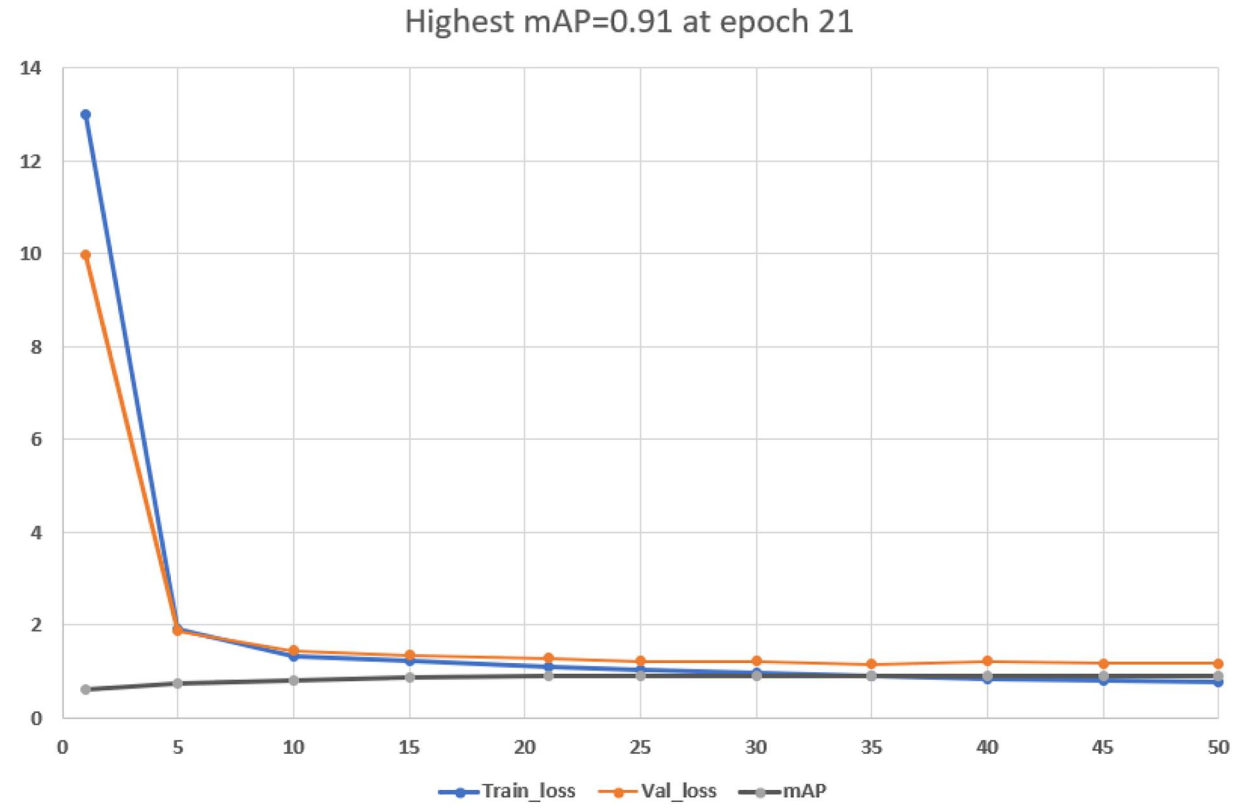

FIGURE 13 Graph of train loss, validation loss and mean average precision (mAP) of the overall experiment reported until 50 epochs. Highest mAP is recorded for epoch 21

TABLE 2 Metrics of model performance

Model	Mean average Precision (%)	Precision (%)	Recall (%)	F1 score (%)	Classification accuracy (%)
ResNet50	88	91	85	89	96
ResNet101	91	96.9	88	92	98

TABLE 3 Comparison of batch sizes against training and validation losses

Batch size	Epochs	Train loss	Validation loss
1	1	13.014	9.9673
	5	1.9152	1.8796
	10	1.3307	1.4471
	15	1.216	1.3613
	20	1.0977	1.2895
	25	1.0229	1.2259
	30	0.9644	1.2187
	35	0.9103	1.1546
	40	0.8516	1.2176
	45	0.8219	1.1718
	50	0.7823	1.1811
5	1	12.775	11.29
	5	4.9746	4.469
	10	2.7309	3.0899
	15	1.9974	2.0535
	20	1.6909	1.6614
	25	1.4612	1.5781
	30	1.3665	1.5517
	35	1.3693	1.3778
	40	1.2955	1.4525
	45	1.2143	1.3676
	50	1.1865	1.3668
10	1	11.519	10.633
	5	1.8724	1.9503
	10	1.3511	1.4069
	15	1.1707	1.2847
	20	1.1045	1.2601
	25	1.0186	1.1916
	30	0.969	1.2302
	35	0.9195	1.1932
	40	0.8686	1.1225
	45	0.8296	1.1439
	50	0.7795	1.1556

approaches [35]. A summary of the data, model, score, and performance is reported in Table 4. As discussed earlier, the performance score of the algorithm differs depending upon the task. So, the results are compared considering the ones that have used a similar performance. Milioto et al., [32] have reported accuracy of 84.62% on classification tasks while using a hybrid of PCA, logistic regression, and auto-encoder. Luus et al., [29] have reported an accuracy of 93.48% on classification task while using CNNs. Sorensen et al. [33] have reported an accuracy of 97% on classification task while using DenseNet. Saldana et al., [34] have reported an 80% localization accuracy, 97.5% classification accuracy, and 0.89 of F1 score on localization and segmentation task while using an adapted version of YOLO and SegNet. In this work, 96% classification accuracy was achieved with ResNet50% and 98% classification accuracy achieved with ResNet101. F1 score of resnet50 is 89% and using ResNet101 it is 92%. It is worth mentioning that all the aforementioned experiments except (Saldana et al., [34]) dealt only with the classification task. An approach is proposed which not only performs classification but also locates coconut trees and segments the trees. Evaluation of an additional performance metric is carried out, the mean Average Precision (mAP). This study achieved 88% mAP using ResNet50% and 91% mAP with backbone architecture ResNet101. This metric shows how accurate the model is to locate and classify the coconut trees.

6 | CONCLUSION AND FUTURE DIRECTION

In this article, an approach is presented for coconut trees' detection and segmentation in aerial imagery of the kingdom of Tonga (South Pacific Islands). A Mask R-CNN-based model using ResNet50 and ResNet101 backbone architectures is reported. The model is trained on the data which is processed and prepared from a single high-resolution aerial image along with the shape file. Experimental results have shown that the proposed model is able to predict coconut trees with a high accuracy (91% mean average precision). This model can be effortlessly extended to classify and locate other kinds of food trees as well. A comparative setup showed that better accuracy is obtained for the ResNet101 architecture when compared with the performance of a ResNet50- based model. Moreover, it carries the benefits of faster R-CNN which is faster than conventional R-CNN and more accurate than CNN. The work carries significance in food resource assessment, humanitarian aid services, and damage analysis in disaster-hit areas, using high-resolution satellite imagery.

The research work is one of the attempts to classify and locate coconut trees based on remote sensed aerial imagery dataset. There is much more potential for future studies in this area. One task of particular significance is to get a cleaner dataset and have methods to get better annotations as these will improve the model training. Future task includes model development to detect other types of food trees (mango, banana, papaya), as well as road conditions and their types.

TABLE 4 Comparison with other techniques

Reference	Data	Model	Score
Chen et al. (2014) [28]	Hyperspectral imagery	Hybrid of PCA, logistic regression, and auto-encoder	Pavia: 84.62%(CA) 0.8451% (F1)
Luus et al. (2015)[29],	Aerial imagery	Author defined CNN	93.48% (CA)
Mortensen et al. (2016) [30],	Photograph by sonny a7	Adapted version of VGG16	79%(CA), 0.66% (IOU)
Lu et al. (2017) [31],	UAV imagery	Author defined CNN	89.5% (CA)
Milioto et al. (2017)[32],	UAV imagery	Author defined CNN	97.50% (CA)
Sorensen et al. (2017) [33],	Photograph by Canon PowerShot G15	DenseNet	97.0% (CA)
Saldana et al. (2019) [34],	Aerial imagery	Adapted version of SegNet and YOLO	97.5% (CA),0.89 (F1)
This article,	UAV aerial imagery	Adapted version of Mask RCNN with backbone ResNet50	96% (CA), 89% (F1), 88% (mAP)
This article	UAV aerial imagery	Adapted version of Mask RCNN with backbone ResNet101	98%(CA), 92% (F1), 91% (mAP)

Abbreviation: IOU, intersection over union; PCA, Principal component analysis; RCNN, region-based convolution neural network; UAV, unmanned air vehicle; VGG, VGG.

ORCID

Hazrat Ali https://orcid.org/0000-0003-3058-5794

ENDNOTES

[1] https://www.unisdr.org/

[2] https://www.openstreetmap.org

[3] https://gdal.org/

[4] https://pypi.org/project/Fiona/

[5] http://www.robots.ox.ac.uk/vgg/software/via/

[6] All the experiments are performed using Intel(R) Core i5-7300HQ CPU 2.50 GHz (4 logical processors), with NVIDIA GeForce GTX 1050 4.00 GB memory and 8 GB of RAM on a 64-bit operating system.

REFERENCES

1. Fritz, A.: Tropical cyclone gita is a monster category 4, and it's hammering Tonga. In: The Washington Post. (The Washington Post (2018). https://www.washingtonpost.com/. Accessed 14 May 2019
2. Assembly, U.N.G.: International strategy for disaster reduction. A/RES/70/204. United Nations-General Assembly (2015)
3. Leetaru, K.: Using ai for good: a new data challenge to use ai to triage natural disaster aerial imagery. In: Forbes. Forbes. (2018). https://www.forbes.com. Accessed 14 May 2019
4. Bank, W. World bank: Automated feature detection of aerial imagery from south pacific - live - google docs. Google Docs (2018). Accessed 14 May 2019
5. Yang, M.Y., et al.: Deep learning for vehicle detection in aerial images. In: 2018 25th IEEE International Conference on Image Processing (ICIP), 3079–3083.IEEE, Athens (2018)
6. Sommer, L.W., Schuchert, T., Beyerer, J.: Fast deep vehicle detection in aerial images. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 311–319.IEEE, Santa Rosa (2017)
7. Luque, B., et al.: Spatio-temporal road detection from aerial imagery using cnns. In: Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 4, 493–500.VISAPP.SCITEPRESS, Porto (2017)
8. Dalponte, M., Bruzzone, L., Gianelle, D.: Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas. IEEE Trans. Geosci. Rem. Sens. 46(5), 1416–1427 (2008)
9. Frolking, S., et al.: Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China. Global Biogeochem. Cycles. 16(4), 38–1 (2002)
10. Rhee, J., Im, J., Carbone, G.J.: Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. Remote Sens. Environ. 114(12), 2875–2887 (2010)
11. Li, Z., Itti, L.: Saliency and gist features for target detection in satellite images. IEEE Trans. Image Process. 20(7), 2017–2029 (2010)
12. Campbell, J.B., Wynne, R.H.: Introduction to remote sensing. Guilford Press, New York (2011)
13. Schmidhuber, J: Deep learning in neural networks: an overview. Neural Network. 61, 85–117 (2015)
14. Rusk, N.: Deep learning. Nat. Methods. 13(1), 35 (2015)
15. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22(10), 1345–1359 (2009)
16. Demmers, T.G., et al.: Simultaneous monitoring and control of pig growth and ammonia emissions. In: 2012 IX International Livestock environment Symposium (ILES IX), 3. American Society of Agricultural and Biological Engineers, Valencia (2012)
17. Sehgal, G., et al.: Crop planning using stochastic visual optimization. In: 2017 IEEE Visualization in Data Science (VDS), 47–51.IEEE, Phoenix (2017)
18. Amara, J., et al.: A deep learning-based approach for banana leaf diseases classification. In: BTW (Workshops). 79–88 Gesellschaft für Informatik, Bonn (2017)
19. Zhao, Z.Q., et al.: Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learn. Syst. 30, 3212–3232 (2019).
20. Girshick, R., et al.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 38(1), 142–158 (2015)
21. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, 1440–1448 IEEE, Santiago (2015)
22. Ren, S., et al.: Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, 91–99 (2015)
23. He, K., et al.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, 2961–2969 (2017)
24. Maghsoudi, O.H., et al.: Automatic breast segmentation in digital mammography using a convolutional neural network. In: 15th International Workshop on Breast Imaging (IWBI2020), 11513, 1151322. Society of Photo-Optical Instrumentation Engineers SPIE, Leuven (2020)
25. Liang, J., et al.: Deep polygon transformer for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9131–9140 (IEEE, Las Vegas 2020)
26. Keskar, N.S., Socher, R.: Improving generalisation performance by switching from adam to sgd. ArXiv (2017). abs/1712.07628
27. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition, 770–778 IEEE, Las Vegas (2016)

28. Chen, Y., Zhao, X., Jia, X.: Spectral–spatial classification of hyperspectral data based on deep belief network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8(6), 2381–2392 (2015)
29. Luus, F.P., et al.: Multiview deep learning for land-use classification. IEEE Geosci. Rem. Sens. Lett. 12(12), 2448–2452 (2015)
30. Mortensen, A.K., et al.: Semantic segmentation of mixed crops using deep convolutional neural network. In: CIGR-AgEng Conference, 26-29 June 2016, Aarhus. Abstracts and Full papers, 1–6. Organising Committee, CIGR (2016)
31. Lu, H., et al.: Cultivated land information extraction in uav imagery based on deep convolutional neural network and transfer learning. J. Mt. Sci. 14(4), 731–741 (2017)
32. Milioto, A., Lottes, P., Stachniss, C.: Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 4, 41 (2017)
33. Jonquet, C., et al.: Agroportal: a vocabulary and ontology repository for agronomy. Comput. Electron. Agric. 144, 126–143 (2018)
34. Ochoa, K.S., Guo, Z.: A framework for the management of agricultural resources with automated aerial imagery detection. Comput. Electron. Agric. 162, 53–69 (2019)
35. Kamilaris, A., Prenafeta.Boldú, F.X.: Deep learning in agriculture: a survey. Comput. Electron. Agric. 147, 70–90 (2018)

How to cite this article: Iqbal, M.S., et al.: Coconut trees detection and segmentation in aerial imagery using mask region-based convolution neural network. IET Comput. Vis. 15(6), 428–439 (2021). https://doi.org/10.1049/cvi2.12028